Ph.D. Title Proposal

On

Enhancing Security Through Resource-Efficient Adversarial Attack Detection using Deep Learning

Submitted in the partial fulfilment of the requirement

for the award of the Degree of

Doctor of Philosophy in Computer Applications

By

Miss. Harshada Umesh Salvi

Under the Guidance of

Dr. Priya Chandran

Computer Applications

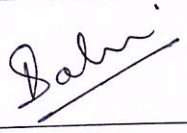
Bharati Vidyapeeth's

Institute of Management and Information Technology

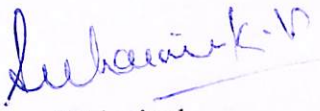
CBD Belapur, Navi Mumbai -400614

University of Mumbai

PROPOSAL FOR TOPIC APPROVAL TO BE SUBMITTED TO THE UNIVERSITY OF MUMBAI FOR THE PH.D. DEGREE IN THE SUBJECT OF COMPUTER APPLICATIONS

Title of the topic	:	Enhancing Security Through Resource-Efficient Adversarial Attack Detection using Deep Learning
Name of the student	:	Harshada Umesh Salvi
Basic Qualification	:	MCA
In the case of Ph.D. student Year of Passing of PET/NET/SET/M.Phil /GPAT/GATE (Teaching Experience)	:	December 2024
Name of the Research Supervisor/Guide	:	Dr. Priya Chandran
Date of Superannuation/Retirement of the Research Supervisor/Guide	:	31/5/2038
Name of the Research Centre	:	Bharati Vidyapeeth's Institute of Management and Information Technology, CBD Belapur, Navi Mumbai
Signature of the Student	:	
Signature of the Research Supervisor/Guide	:	

Date: 14/8/2025



Signature of Principal

PRINCIPAL
Bharati Vidyapeeth's
Institute of Management & Information Technology
Sector-8, C.B.D. Belapur,
Navi Mumbai-400 614

Declaration

I wish to state that the work embodied in this synopsis titled “Enhancing Security Through Resource-Efficient Adversarial Attack Detection using Deep Learning” forms my own contribution to the work carried out under the guidance of Dr. Priya Chandran at the Bharti Vidyapeeth’s Institute of Management and Information Technology. I declare that this written submission represents my ideas in my own words and where others’ ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission.

Miss. Harshada U. Salvi

Table of Contents

Sr. No	Topic Name
01	Introduction
02	A survey of Work Done in the Research Area
03	Gap in Literature Survey
04	Rationale and Significance of Study
05	Aims and Objectives of the Study
06	Methodologies and Techniques to be used
07	Plan of Research
08	Expected Outcomes of the study
09	References

Research Title: Enhancing Security Through Resource-Efficient Adversarial Attack Detection using Deep Learning

1. Introduction

The proliferation of resource-constrained devices such as Internet of Things (IoT) sensors, wearable technology, and embedded systems has brought new possibilities across various domains, including smart homes, healthcare, agriculture, and industrial automation. As these devices collect and transmit sensitive data and perform critical tasks autonomously, ensuring their security and integrity has become a top priority.

However, the increasing deployment of resource-constrained devices brings forth new security threats, especially adversarial attacks that involve subtle, often imperceptible manipulations of input data that can mislead even state-of-the-art deep learning (DL) models into making incorrect decisions. These attacks can compromise system integrity, privacy, and safety, making robust and efficient detection mechanisms crucial.

Deep learning has become a core component in modern intrusion detection and classification systems due to its remarkable ability to learn complex patterns from large-scale data. However, Deep Learning models are inherently vulnerable to adversarial perturbations. Given the computational and energy constraints of resource-constrained devices, the development of lightweight yet accurate adversarial attack detection mechanisms become a research priority.

By leveraging recent advancements in deep learning architectures this research seeks to design and validate a framework that can detect adversarial threats accurately and efficiently.

2. A survey of Work Done in the Research Area

The security landscape of IoT is vast and evolving, with foundational overviews such as El Hajla et al. [1] outlining major challenges such as scalability, heterogeneity, and real-time security enforcement. Complementing this, Srivastava et al. [2] delve into future vulnerabilities, stressing that the explosive growth of connected devices increases the attack surface and necessitates robust threat modeling frameworks. Golandaz and Sharma [3] discussed how IoT devices are being misused, such as being turned into botnets or spying tools. Msgna [4] provides a comprehensive anatomy of common attack vectors in IoT—including DoS, spoofing, and side-channel attacks—and the efficacy of layered mitigation strategies. Lavanya [5] highlighted how IoT devices often lack basic protection and stressed the need for standard rules, device-level authentication, and regular software updates.

Hussain et al. [6] reviewed how machine learning (ML) is being used to secure IoT, including models that can detect unusual behaviour or intrusions. However, they also noted issues like lack of data and vulnerabilities to adversarial attacks. Praveena et al. [7] focused on these adversarial attacks—small changes made to fool ML models—and categorized different types, such as poisoning and evasion attacks. They stressed the need to train models to resist such attacks. Khazane et al. [8] took a broader look at these threats, suggesting we need to develop more transparent and robust ML models for IoT.

Harbi et al. [9] proposed a roadmap to improve IoT security using adversarial machine learning techniques. They suggested that ML models should be designed with security in mind from the start. Son et al. [10] looked at how 6G networks can support IoT but warned that the increased speed and connectivity could also increase the chances of attacks, calling for faster and more efficient defence mechanisms. Gungor et al. [11] tested different ML models for intrusion detection and found that adversarial attacks could reduce their accuracy significantly, pointing to the need for better evaluation methods.

Trivedi et al. [12] studied the effects of adversarial attacks in industrial systems that use IoT. They found that these attacks could seriously affect decision-making and suggested the use of smarter, context-aware defences. Alkadi et al. [13] reviewed

many uses of adversarial machine learning in IoT and proposed combining multiple learning techniques for better security. Kotak and Elovici [14] showed how even small changes in data can cause device misclassification, which is dangerous for systems that rely on identifying devices accurately.

Mbow et al. [15] demonstrated how generative adversarial networks (GANs) can create fake network traffic that avoids detection. Their study showed how easy it is to trick IoT intrusion detection systems. Qiu et al. [16] compared different deep learning models and found that many of them failed when faced with adversarial inputs. Che et al. [17] supported this by showing that even small changes in input data could cause major drops in accuracy across various models.

The DODEM framework [18] a two-layer defence system that combines adversarial training and model checks to make industrial IoT analytics more secure. Lastly, Sharipuddin and Winanto [19] showed that using Deep Belief Networks can make IoT detection systems more resistant to adversarial attacks while still being lightweight enough for low-resource devices.

3. Gap in Literature Survey

Despite growing interest in adversarial machine learning (AML) in IoT systems, current defence mechanisms remain insufficiently adapted to the unique constraints of IoT environments, which are typically resource-limited, heterogeneous, and latency-sensitive [2], [6], [8], [13]. There is a critical need to develop robust, lightweight, and adaptive AML defences that ensure security without compromising the performance or communication integrity of IoT devices [9], [10], [14]. Furthermore, deep learning-based intrusion detection systems (DL-NIDSs) in IoT remain vulnerable to adversarial inputs, and limited research has addressed how to effectively harden these systems against such threats [16], [17].

Much of the existing literature focuses narrowly on well-known attack techniques like the Fast Gradient Sign Method (FGSM), neglecting more complex or adaptive adversarial methods such as Projected Gradient Descent (PGD), Carlini-Wagner (CW), and GAN-based attacks [7], [15], [16]. In addition, there is limited exploration of adversarial threats in emerging IoT technologies, such as reconfigurable intelligent surfaces and semantic communication systems, where the attack surface and vulnerability patterns may significantly differ [10]. The domain of network-level adversarial attacks within IoT is particularly underexplored, representing a critical gap in the literature [8], [12].

Another key limitation is the lack of comprehensive evaluation of existing defence strategies. Many studies fail to assess how these mechanisms perform across diverse adversarial scenarios or under real-world operational constraints [6], [11]. Moreover, the performance trade-offs—such as computational cost versus security gain—are rarely analysed in detail [17], [18]. Evaluations often lack generalizability, leaving questions about the robustness of current approaches under evolving threat models [8], [11].

Additionally, the integration of hybrid and advanced machine learning techniques for adversarial defence remains underdeveloped. Ensemble models, Deep Belief Networks (DBNs), and context-aware ML approaches could offer improved resilience but are rarely investigated in current research [13], [18], [19]. This reflects a broader

gap in applying adaptive and intelligent strategies to strengthen IoT security frameworks [9].

Finally, few studies consider how adversarial attacks and defences might evolve with the adoption of future technologies such as 6G, intelligent surfaces, or semantic-aware networking. These technologies are expected to dramatically change the IoT communication paradigm, yet their impact on AML threat landscapes and countermeasure effectiveness remains largely unaddressed [3], [5], [10]. This highlights a pressing need for forward-looking studies that anticipate future vulnerabilities and develop proactive defence strategies accordingly.

4. Rationale and Significance of Study

As the use of resource-constrained devices grows rapidly in areas like healthcare, smart homes, agriculture, and industries, ensuring that these devices are secure has become extremely important. Traditional rule-based security systems and shallow machine learning models often fail to recognize sophisticated adversarial patterns, especially those designed to subtly manipulate input data. Deep learning models, with their ability to automatically learn complex, high-dimensional feature representations, are uniquely suited for capturing subtle anomalies that adversarial examples introduce often invisible to humans or simpler algorithms.

Adversarial machine learning has emerged as a critical threat vector in the AI-driven resource-constrained landscape. Most existing defence mechanisms are computationally expensive and tailored for powerful computing environments, making them unsuitable for resource-constrained systems. The significance of this study lies in designing a resource-efficient deep learning models that can efficiently detect and mitigate adversarial inputs without overwhelming the limited resources of IoT devices. A successful framework could drastically enhance the resilience and trustworthiness of smart environments, protect critical infrastructures, and encourage broader adoption of AI-enabled resource-constrained applications.

5. Statement of Aim and Objectives

Aim:

To develop a resource-efficient deep learning-based adversarial attack detection framework that enhances the security of systems operating in resource-constrained environments.

Research Objectives

1. To study and analyse existing adversarial attack patterns and detection mechanisms suitable for resource constrained environments.
2. To design an intelligent and resource-efficient detection system using deep learning techniques that can effectively identify adversarial or manipulated inputs to enhance system security.
3. To evaluate the performance of the proposed system under various attack scenarios in order to assess its robustness, reliability, and suitability for resource constraint environments.
4. To compare the proposed system with state-of-the-art detection mechanisms.

6. Methodologies and Techniques to be Used

The literature study involves a comprehensive review of existing research to understand various adversarial attack techniques, detection strategies using deep learning, models suitable for resource-constrained environments, and the specific security challenges faced by such systems. In the implementation phase, appropriate datasets will be identified and collected to ensure meaningful and comparable results. A resource-efficient deep learning model will then be designed, developed, and trained to detect adversarial attacks effectively. This phase also includes the generation of adversarial examples using attack generation methods and the application of model optimization techniques to minimize model size and computational requirements making them suitable for deployment on resource-constrained devices. The final phase, assessment, focuses on evaluating the performance of the developed model under simulated conditions. This evaluation will measure the model's reliability, efficiency, and suitability for deployment in resource-constrained environments. Additionally, the proposed model will be compared with existing state of the art models to benchmark performance, and its robustness will be tested against various adversarial attack scenarios. The outcomes of these evaluations will be thoroughly analysed and discussed to validate the effectiveness of the proposed approach.

Proposed Architecture

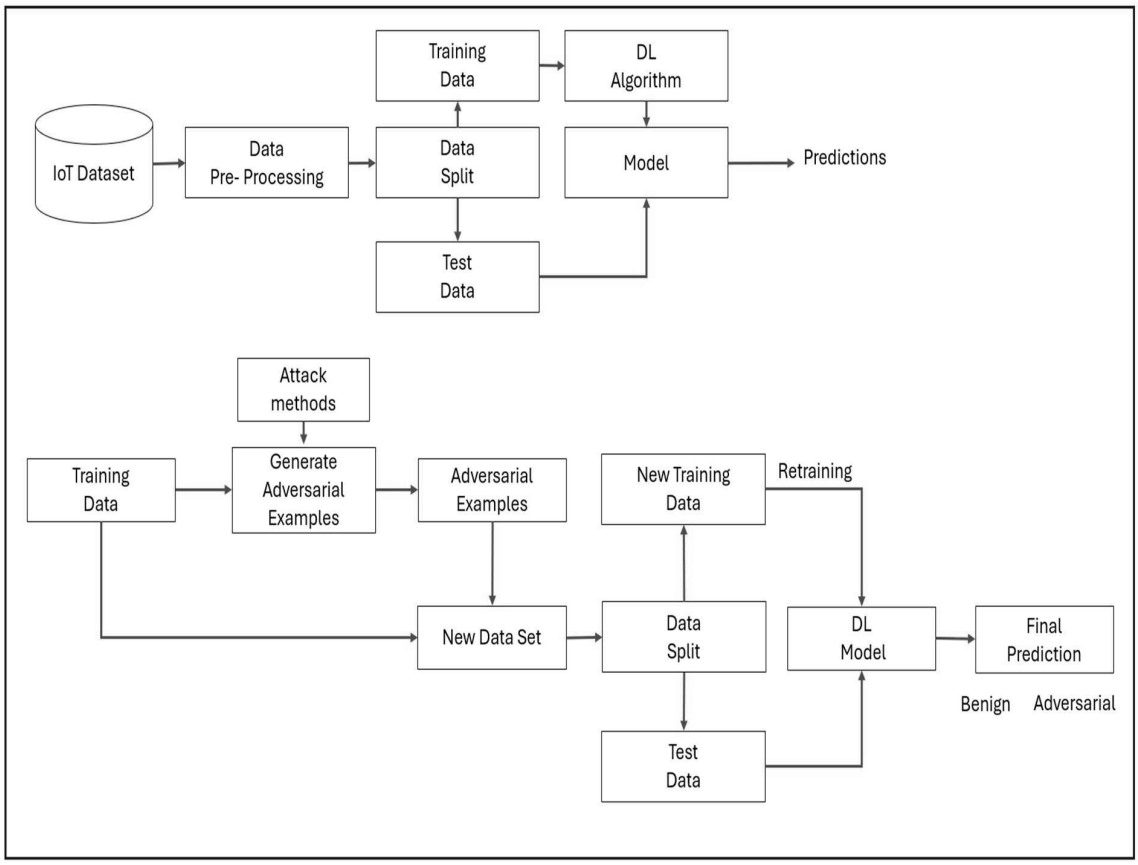


Figure 1 Proposed Architecture

Step 1 – Dataset Selection & Preparation - Relevant datasets related to IoT will be chosen. Data will be cleaned, normalized, and formatted for training deep learning models.

Step 2 – Design resource-efficient Deep Learning Model - Develop a deep learning model tailored for resource constrained environments to classify clean data. The model will be trained on clean data.

Step 3 – Generate Adversarial Examples - Use adversarial attack techniques to craft inputs that can fool Deep learning models.

Step 4 - Train and Validate the Detection Model - Retrain the proposed model to distinguish between normal and adversarial inputs using clean and perturbed data. The performance of the proposed model will be validated.

Step 5 - Evaluate Robustness Against Different Attack Scenarios - Test the model's strength against multiple attack methods and measure how well it detects them.

Step 6 - Analyse Results and Compare with existing model - Compare proposed model's performance with state of the art models in terms of detection accuracy and resource usage.

7. Plan of Research

Sr. No	Activity	Duration (Each block of 3 months)
1	Extensive Literature Review and Problem Statement	
2	Analysis of existing adversarial attack patterns and detection mechanisms suitable for resource constrained environments.	
3	Design and implement resource-efficient deep learning model to detect maliciously crafted input.	
4	Evaluate the performance of the developed model under different attack scenarios	
5	Compare the proposed model with existing state of the art deep learning model.	
6	Analyse and discuss results, Documentation	

8. Expected Outcomes of Study

The study is expected to result in the development of a resource-efficient, deep learning-based adversarial detection technique that is both accurate and efficient enough to operate within the computational and memory constraints. This model will effectively differentiate between benign and adversarial inputs in real time, addressing a critical security gap in smart environments.

The research will also produce comparative performance results that highlight the advantages of the proposed lightweight technique over traditional adversarial defence mechanisms.

The successful implementation of this framework can significantly enhance the security of resource constrained systems. Ensuring that these systems are protected against adversarial manipulation will help maintain trust, safety, and reliability.

Moreover, the study will lay a strong academic and industrial foundation for the standardization of adversarial detection mechanisms, enabling scalable and secure AI adoption across diverse resource-constraint ecosystems.

9. References

[1] S. El Hajla, E. M. Ennaji, Y. Maleh, and S. Mounir, "Security Challenges and Solutions in IoT," *Advances in Computational Intelligence and Robotics*, pp. 25–50, Nov. 2024, doi: 10.4018/979-8-3693-7540-2.ch002.

[2] A. Srivastava, S. Gupta, M. Quamara, P. Chaudhary, and V. Aski, "Future IoT-enabled threats and vulnerabilities: State of the art, challenges, and future prospects," *Int. J. Commun. Syst.*, vol. 33, no. 12, Aug. 2020, doi: 10.1002/dac.4443.

[3] A. Golandaz and U. Sharma, "IoT Under Siege: The Dark Side of Internet-Connected Devices," *Int. J. Multidiscip. Res.*, vol. 6, no. 3, 2024, doi: 10.36948/ijfmr.2024.v06i03.22797.

[4] M. Msgna, "Anatomy of attacks on IoT systems: review of attacks, impacts and countermeasures," *J. Surveillance, Security and Safety*, vol. 3, no. 4, pp. 150–173, Jan. 2022, doi: 10.20517/jsss.2022.07.

[5] A. Lavanya, "Safeguarding IoT Devices Against Emerging Security Threats: Challenges and Mitigation Strategies," *Int. J. Sci. Res.*, vol. 13, no. 12, pp. 448–453, Dec. 2024, doi: 10.21275/sr241205115818.

[6] F. Hussain, R. Hussain, S. A. Hassan, and E. Hossain, "Machine Learning in IoT Security: Current Solutions and Future Challenges," *IEEE Commun. Surv. Tutor.*, vol. 22, no. 3, pp. 1686–1721, 2020, doi: 10.1109/COMST.2020.2986444.

[7] M. Praveena, S. Madhumitha, J. Menakadevi, and V. L. Akkash, "A Comprehensive Taxonomy of Adversarial Attacks on Machine Learning in IoT Application," in *Proc. ICECA*, 2023, pp. 1398–1403, doi: 10.1109/ICECA58529.2023.10394928.

[8] H. Khazane, M. Ridouani, F. Salahdine, and N. Kaabouch, "A Holistic Review of Machine Learning Adversarial Attacks in IoT Networks," *Future Internet*, vol. 16, no. 1, 2024, doi: 10.3390/fi16010032.

[9] Y. Harbi, K. Medani, C. Gherbi, Z. Aliouat, and S. Harous, "Roadmap of Adversarial Machine Learning in Internet of Things-Enabled Security Systems," *Sensors*, vol. 24, no. 16, 2024, doi: 10.3390/s24165150.

[10] B. D. Son *et al.*, "Adversarial Attacks and Defenses in 6G Network-Assisted IoT Systems," *arXiv preprint arXiv:2401.14780*, Jan. 2024, doi: 10.48550/arxiv.2401.14780.

[11] O. Gungor *et al.*, "Rigorous Evaluation of Machine Learning-Based Intrusion Detection Against Adversarial Attacks," in *Proc. IEEE Int. Conf. Cyber Secur. Resilience (CSR)*, London, U.K., 2024, pp. 152–158, doi: 10.1109/CSR61664.2024.10679443.

[12] S. Trivedi, N. Faruqui, and M. A. Hassan, "An Exploratory Analysis of Effect of Adversarial Machine Learning Attack on IoT-enabled Industrial Control Systems," in *Proc. Int. Conf. Softw. Comput. Appl.*, 2023, pp. 1–8, doi: 10.1109/ICSCA57840.2023.10087713.

[13] S. H. Alkadi, S. Al-Ahmadi, and M. M. Ben Ismail, "Better Safe Than Never: A Survey on Adversarial Machine Learning Applications towards IoT Environment," *Appl. Sci.*, vol. 13, no. 10, p. 6001, 2023, doi: 10.3390/app13106001.

[14] J. Kotak and Y. Elovici, "Adversarial Attacks Against IoT Identification Systems," *IEEE Internet Things J.*, vol. 10, no. 9, pp. 7868–7883, May 2023, doi: 10.1109/JIOT.2022.3229906.

[15] M. Mbow, R. Roman, T. Takahashi, and K. Sakurai, "Evading IoT Intrusion Detection Systems with GAN," in *Proc. AsiaJCIS*, 2024, pp. 48–55, doi: 10.1109/ASIAJCIS64263.2024.00018.

[16] H. Qiu, T. Dong, T. Zhang, J. Lu, G. Memmi, and M. Qiu, "Adversarial Attacks Against Network Intrusion Detection in IoT Systems," *IEEE Internet Things J.*, vol. 8, no. 13, pp. 10327–10335, Jul. 2021, doi: 10.1109/JIOT.2020.3048038.

[17] X. Che, H. Li, L. Chi, F. Wang, and J. Hu, "Evaluation of the Robustness of IoT Intrusion Detection Models in Adversarial Attack Environments," SSRN, 2023, doi: 10.2139/ssrn.4611920.

[18] Gungor, O., Rosing, T. and Aksanli, B. (2023) DODEM: Double defense mechanism against adversarial attacks towards Secure Industrial Internet of Things Analytics, arXiv.org. Available at: https://arxiv.org/abs/2301.09740.

[19] S. Sharipuddin and E. A. Winanto, "Defence against adversarial attacks on IoT detection systems using Deep Belief Network," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 35, no. 2, p. 1073, 2024, doi: 10.11591/ijeecs.v35.i2.pp1073-1081.